



PATENT
Attorney Docket No. 3495.0010-01

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:)
)
Marc Alizon, et al.)
)
Serial No.: 07/158,652) Group Art Unit: 1805
)
Filed: February 22, 1988) Examiner: J. Railey
)
For: CLONED DNA SEQUENCES RELATED)
TO THE GENOMIC RNA OF)
LYMPHADENOPATHY ASSOCIATED)
VIRUS (LAV) AND PROTEINS)
ENCODED BY SAID LAV GENOMIC)
RNA)

Honorable Commissioner of Patents
and Trademarks
Washington, D.C. 20231

Sir:

CLAIM FOR PRIORITY

Under the provisions of Section 119 of 35 U.S.C.,
applicants hereby claim the benefit of the filing date of Great
Britain Application No. 84 29099, filed November 16, 1984, for
the above identified United States Patent Application.

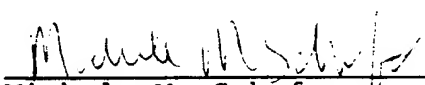
In support of applicants' claim for priority, filed
herewith is one certified copy of GB 84 29099.

If there are any fees due in connection with the filing of this Paper, please charge such fees to our Deposit Account No. 06-0916.

Respectfully submitted,

FINNEGAN, HENDERSON, FARABOW,
GARRETT & DUNNER

By:


Michele M. Schäfer
Reg. No. 34,717

Date: October 21, 1993

LAW OFFICES
FINNEGAN, HENDERSON
FARABOW, GARRETT
& DUNNER
1300 I STREET, N.W.
WASHINGTON, DC 20005
1-202-406-4000



THE PATENT OFFICE,
25 SOUTHAMPTON BUILDINGS,
LONDON.

I, the undersigned, being an officer duly authorised in accordance with Section 62(3) of the Patents and Designs Act 1907, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents, has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or the inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Please turn over

Re-registration under the Companies Act does not constitute a new legal entity
but merely subjects the company to certain additional company law rules.

29th Witness my hand this
day of OCTOBER 19 85

COC.1

W. L. L.

PATENTS ACT 1977

16 NOV 1984

PATENTS FORM No. 1/77 (Revised 1982)
(Rules 16, 19)

The Comptroller
The Patent Office
25 Southampton Buildings
London, WC2A 1AY

19/11/84 B3662 PAT*** 10.0

1984

29099

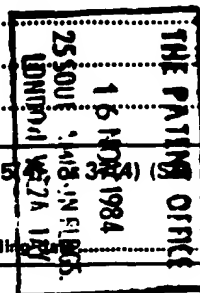
Fee: £10.00

REQUEST FOR GRANT OF A PATENT

8429099

THE GRANT OF A PATENT IS REQUESTED BY THE UNDERSIGNED ON THE BASIS OF THE PRESENT APPLICATION

I	Agent's Reference	JJD/EAF/26804
II	Title of Invention	<u>CLONED DNA SEQUENCES RELATED TO THE GENOMIC RNA OF LYMPHADENOPATHY-ASSOCIATED VIRUS (LAV) AND PROTEINS ENCODED BY SAID LAV GENOMIC RNA.</u>
III	Applicant or Applicants (See note 2)	
	Name (First or only applicant)	<u>INSTITUT PASTEUR</u>
	Country	<u>FRANCE</u>
	State	
	ADP Code No.	
	Address	<u>25-28 Rue du Dr. Roux,</u> <u>75724 Paris Cedex 15, France.</u>
	Name (of second applicant, if more than one)	<u>Centre National</u>
	<u>de la Recherche Scientifique</u>	
	Country	<u>FRANCE</u>
	State	
	Address	<u>15 Quai Anatole France</u> <u>75007 Paris, France.</u>
IV	Inventor (see note 3)	<u>Is/are represented by the undersigned</u> or (b) A statement on Patents Form No. 7/77 is/will be furnished
V	Name of Agent (if any) (See note 4)	<u>Reddie & Grose</u>
	ADP CODE NO	
VI	Address for Service (See note 5)	<u>16 Theobalds Road</u> <u>London WC1X 8PL</u>
VII	Declaration of Priority (See note 6)	
	Country	
	Filing date	
	File number	
VIII	The Application claims an earlier date under Section 8(3), 12(6), 15(4) or 31(4) (See note 7)	
	Section No.	
	Earlier application or patent number	
	and filing date	



IX Check List (To be filled in by applicant or agent)

- | | |
|---|---|
| A The application contains the following number of sheet(s) | B The application as filed is accompanied by:- |
| 1 Request 1 Sheet(s) | 1 Priority document No |
| 2 Description 17 Sheet(s) | 2 Translation of priority document No |
| 3 Claim(s) 2 Sheet(s) | 3 Request for Search No |
| 4 Drawing(s) 26 Sheet(s) | 4 Statement of Inventorship and Right to Apply No |
| 5 Abstract 0 Sheet(s) | 5 |

X It is suggested that Figure No 1 of the drawings (if any) should accompany the abstract when published.

XI Signature (See note 8)


Reddie & Grose, Agents for the Applicant(s)

NOTES:

1. This form, when completed, should be brought or sent to the Patent Office together with the prescribed fee and two copies of the description of the invention, and of any drawings.
2. Enter the name and address of each applicant. Names of individuals should be indicated in full and the surname or family name should be underlined. The names of all partners in a firm must be given in full. Bodies corporate should be designated by their corporate name and the country of incorporation and, where appropriate, the state of incorporation within that country should be entered where provided. Full corporate details, eg "a corporation organised and existing under the laws of the State of Delaware, United States of America," trading styles, eg "trading as xyz company", nationality, and former names, eg "formerly [known as] ABC Ltd," are not required and should not be given. Also enter applicant(s) ADP Code No. (if known).
3. Where the applicant or applicants is/are the sole inventor or the joint inventors, the declaration (a) to that effect at IV should be completed, and the alternative statement (b) deleted. If, however, this is not the case the declaration (a) should be struck out and a statement will then be required to be filed upon Patent Form No 7/77.
4. If the applicant has appointed an agent to act on his behalf, the agent's name and the address of his place of business should be indicated in the spaces available at V and VI. Also insert agent's ADP Code No. (if known) in the box provided.
5. An address for service in the United Kingdom to which all documents may be sent must be stated at VI. It is recommended that a telephone number be provided if an agent is not appointed.
6. The declaration of priority at VII should state the date of the previous filing and the country in which it was made and indicate the file number, if available.
7. When an application is made by virtue of section 8(3), 12(8), 15(4), or 37(4) the appropriate section should be identified at VIII and the number of the earlier application or any patent granted thereon identified.
8. Attention is directed to rules 90 and 106 of the Patent Rules 1982.
9. Attention of applicants is drawn to the desirability of avoiding publication of inventions relating to any article, material or device intended or adapted for use in war (Official Secrets Acts, 1911 and 1920). In addition after an application for a patent has been filed at the Patent Office the comptroller will consider whether publication or communication of the invention should be prohibited or restricted under section 22 of the Act and will inform the applicant if such prohibition is necessary.
10. Applicants resident in the United Kingdom are also reminded that, under the provisions of section 23 applications may not be filed abroad without written permission or unless an application has been filed not less than six weeks previously in the United Kingdom for a patent for the same invention and no direction prohibiting publication or communication has been given or any such direction has been received.

Cloned DNA sequences related to the genomic RNA of lymphadenopathy-associated-virus (LAV) and proteins encoded by said LAV genomic RNA

The invention relates to cloned DNA sequences indistinguishable from genomic RNA and DNA of lymphadenopathy-associated virus (LAV), a process for their preparation and their uses. It relates more particularly to stable probes including a DNA sequence which can be used for the detection of the LAV virus or related viruses or DNA proviruses in any medium, particularly biological samples containing any of them. The invention also relates to polypeptides, whether glycosylated or not, encoded by said DNA sequences.

Lymphadenopathy-associated virus (LAV) is a human retrovirus first isolated from the lymph node of a homosexual patient with lymphadenopathy syndrome, frequently a prodrome or a benign form of acquired immune deficiency syndrome (AIDS). Subsequently other LAV isolates have been recovered from patients with AIDS or pre-AIDS. All available data are consistent with the virus being the causative agent of AIDS.

A method for cloning such DNA sequences has already been disclosed in British Patent Application No. 84 23858 filed on September 19, 1984. Reference is hereafter made to that application as concerns subject matter in common with the further improvements to the invention disclosed herein.

The present invention aims at providing additional new means which should not only also be useful for the detection of LAV or related viruses (hereafter more generally referred to as "LAV viruses"), but also have more versatility, particularly in detecting specific parts of the genomic DNA of said viruses whose expression products are not always directly detectable by immunological methods.

The present invention further aims at providing

polypeptides containing sequences in common with polypeptides encoded by the LAV genomic RNA. It relates even more particularly to polypeptides comprising antigenic determinants included in the proteins encoded and expressed by the LAV genome occurring in nature. An additional object of the invention is to further provide means for the detection of proteins related to LAV virus, particularly for the diagnosis of AIDS or pre-AIDS or, to the contrary, for the detection of antibodies against the LAV virus or proteins related therewith, particularly in patients afflicted with AIDS or pre-AIDS or more generally in asymptomatic carriers and in blood-related products. Finally the invention also aims at providing immunogenic polypeptides, and more particularly protective polypeptides for use in the preparation of vaccine compositions against AIDS or related syndromes.

The present invention relates to additional DNA fragments, hybridizable with the genomic RNA of LAV as they will be disclosed hereafter, as well as with additional cDNA variants corresponding to the whole genomes of LAV viruses. It further relates to DNA recombinants containing said DNAs or cDNA fragments.

The invention relates more particularly to a cDNA variant corresponding to the whole of LAV retroviral genomes, which is characterized by a series of restriction sites in the order hereafter (from the 5' end to the 3' end).

The coordinates of the successive sites of the whole LAV genome (restriction map) are indicated hereafter too, with respect to the Hind III site (selected as of coordinate 1) which is located in the R region. The coordinates are estimated with an accuracy of ± 200 bp :

	Hind III	0
	Sac I	50
35	Hind III	520
	Pst I	800
	Hind III	1 100

	Bgl II	1 500
	Kpn I	3 500
	Kpn I	3 900
	Eco RI	4 100
5	Eco RI	5 300
	Sal I	5 500
	Kpn I	6 100
	Bgl II	6 500
	Bgl II	7 600
10	Hind III	7 850
	Bam HI	8 150
	Xho I	8 800
	Kpn I	8 700
	Bgl II	8 750
15	Bgl II	9 150
	Sac I	9 200
	Hind III	9 250

Another DNA variant according to this invention optionally contains an additional Hind III approximately at the 5 550 coordinate.

Reference is further made to fig. 1 which shows a more detailed restriction map of said whole-DNA (λ J19).

An even more detailed nucleotide sequence of a preferred DNA according to the invention is shown in fig. 4-12 hereafter.

The invention further relates to other preferred DNA fragments which will be referred to hereafter.

Additional features of the invention will appear in the course of the non-limitative disclosure of additional features of preferred DNAs of the invention, as well as of preferred polypeptides according to the invention. Reference will further be had to the drawings in which :

- fig. 1 is the restriction map of a complete LAV genome (clone λ J19) ;

- figs. 2 and 3 show diagrammatically parts of the three

possible reading phases of LAV genomic RNA, including the open reading frames (ORF) apparent in each of said reading phases ;

- figs. 4-12 show the successive nucleotidic sequences of a complete LAV genome. The possible peptide sequences in relation to the three possible reading phases related to the nucleotidic sequences shown are also indicated ;
- figs. 13-18 reiterate the sequence of part of the LAV genome containing the genes coding for the envelope proteins, with particular boxed peptidic sequences which corresponds to groups which normally carry glycosyl groups.

The sequencing and determination of sites of particular interest was carried out on a phage recombinant corresponding to AJ19 disclosed in the abovesaid British Patent application Nr. 84 23659. A method for preparing it is disclosed in that application.

The whole recombinant phage DNA of clone AJ19 (disclosed in the earlier application) was sonicated according to the protocol of DEININGER (1983), Analytical Biochem. 129, 218. the DNA was repaired by a Klenow reaction for 12 hours at 16°C. The DNA was electrophoresed through 0.8 % agarose gel and DNA in the size range of 300-800 bp was cut out and electroeluted and precipitated. Resuspended DNA (in 10 mM Tris, pH 8 ; 0.1 mM EDTA) was ligated into M13mp8 RF DNA (cut by the restriction enzyme SmaI and subsequently alkaline phosphated), using T4 DNA- and RNA-ligases (Maniatis T et al (1982) - Molecular cloning - Cold Spring Harbor Laboratory). An *E. coli* strain designated as TGI was used for further study. This strain has the following genotype :

Alac pro, supE, thi.F'tre036, proAB, lacI^q, ZAM15,r⁻

This *E. coli* TGI strain has the peculiarity of enabling recombinants to be recognized easily. The blue colour of the cells transfected with plasmids which did

not recombine with a fragment of LAV DNA is not modified. To the contrary cells transfected by a recombinant plasmid containing a LAV DNA fragment yield white colonies. The technique which was used is disclosed in Gene (1983), 28.

5 101.

This strain was transformed with the ligation mix using the Hanahan method (Hanahan D (1983) J. Mol. Biol. 165, 557). Cells were plated out on tryptone-agarose plate with IPTG and X-gal in soft agarose. White plaques were
10 either picked and screened or screened directly in situ using nitrocellulose filters. Their DNAs were hybridized with nick-translated DNA inserts of pUC18 Hind III subclones of AJ19. this permitted the isolation of the plasmids or subclones of A which are identified in the
15 table hereafter. In relation to this table it should also be noted that the designation of each plasmid is followed by the deposition number of a cell culture of E. coli TGI containing the corresponding plasmid at the "Collection Nationale des Cultures de Micro-organismes" (C.N.C.M.) of
20 the Pasteur Institute in Paris, France. A non-transformed TGI cell line was also deposited at the C.N.C.M. under Nr. I-364. All these deposits took place on November 15, 1984. The sizes of the corresponding inserts derived from the LAV genome have also been indicated.

25



TABLE
Essential features of the recombinant plasmids

5	- pJ19 - 1 plasmid	(I-365)	0.5 kb
	Hind III - Sac I - Hind III		
10	- pJ19 - 17 plasmid	(I-367)	0.6 kb
	Hind III - Pst I - Hind III		
	- pJ19 - 6 plasmid	(I-368)	1.5 kb
15	Hind III (5')		
	Bam HI		
	Xho I		
	Kpn I		
	Bgl II		
20	Sac I (3')		
	Hind III		
	- pJ19-13 plasmid	(I-368)	6.7 kb
25	Hind III (5')		
	Bgl II		
	Kpn I		
	Kpn I		
	Eco RI		
30	Eco RI		
	Sal I		
	Kpn I		
	Bgl II		
	Bgl II		
35	Hind III (3')		

M : méthionine
 W : tryptophan
 F : phenylalanine
 Y : tyrosine
 5 L : leucine
 V : valine
 I : isoleucine
 G : glycine
 Y : thréonine
 10 S : serine
 E : glutamic acid
 D : Aspartic acid
 N : asparagine
 Q : glutamine
 15 P : proline.

The asterik signs "*" correspond to stop codons (i.e. TAA, TAG and TGA).

Starting above the first line of the DNA nucleotidic sequence of fig. 4 the three reading phases
 20 are respectively marked "1", "2", "3", on the left handside of the drawing. The same relative presentation of the three theoretical reading phases is then used all over the successive lines of the LAV nucleotidic sequence.

Figs. 2 and 3 provide a diagrammatized representation of the lengths of the successive open reading
 25 frames corresponding to the successive reading phases (also referred to by numbers "1", "2" and "3" appearing in the left handside part of fig. 2). The relative positions of these open reading frames (ORF) with respect to the
 30 nucleotidic structure of the LAV genome is referred to by the scale of numbers representative of the respective positions of the corresponding nucleotides in the DNA sequence. The vertical bars correspond to the positions of the corresponding stop codons.

35 1) The "gag gene" (or ORF-gag)

The "gag gene" codes for core proteins.

Particularly it appears that a genomic fragment (ORF-gag) thought to code for the core antigens including the p25, p18 and p13 proteins is located between nucleotidic position 238 (starting with 5' CTA GCG GAG 3') and
 5 nucleotidic position 1759 (ending by CTCG TCA CAA 3'). The structure of the peptides or proteins encoded by parts of said ORF is deemed to be that corresponding to phase 2.

The methionine aminoacid "M" coded by the ATG at position 260-262 is the probable initiation methionine of
 10 the gag protein precursor. The end of ORF-gag and accordingly of gag protein appears to be located at position 1759.

The beginning of p25 protein, thought to start by a P-I-V-Q-N-I-Q-G-Q-M-V-M aminoacid sequence is
 15 thought to be coded for by the nucleotidic sequence CCTATA.... starting at position 658.

Hydrophilic peptides in the gag open reading frame are identified hereafter. They are defined starting from aminoacid 1 = Met (M) coded by the ATG starting from 260-2
 20 in the LAV DNA sequence.

Those hydrophilic peptides are
 12-32 aminoacids inclusive

	37-46	"	"
	49-79	"	"
25	88-153	"	"
	158-189	"	"
	178-188	"	"
	200-220	"	"
	226-234	"	"
30	239-264	"	"
	268-331	"	"
	352-381	"	"
	377-390	"	"
	399-432	"	"
35	437-484	"	"
	492-498	"	"

The invention also relates to any combination of these peptides.

2) The "pol gene" (or ORF-pol)

Figs. 4-12 also show that the DNA fragments
5 extending from nucleotidic position 1555 (starting with
5'TTT TTT3' to nucleotidic position 5088 is thought
to correspond to the pol gene. The polypeptidic structure
of the corresponding polypeptides is deemed to be that
corresponding to phase 1. It stops at position 4563 (and
10 by 5'G GAT GAG GAT 3').

These genes are thought to code for the virus
polymerase or reverse transcriptase.

3) The envelope gene (or ORF-env)

The DNA sequence thought to code for envelope
15 proteins is thought to extend from nucleotidic position
5670 (starting with 5'AAA GAG GAG A....3') up to nucleo-
tidic position 8132 (ending byA ACT AAA GAA 3').
Polypeptidic structures of sequences of the envelope
protein correspond to those read according to the "phase
20 3" reading phase.

The start of env transcription is thought to be at
the level of the ATG codon at positions 5681-5693.

Additional feature of the envelope protein coded
by the env genes appear on figs. 13-18. These are to be
25 considered as paired figs. 13 and 14 ; 15 and 16 ; 17 and
18 respectively.

It is to be mentioned that because of format
difficulties.

Fig. 14 overlaps to some extent with fig. 13.

30 Fig. 15 overlaps to some extent with fig. 15.

Fig. 16 overlaps to some extent with fig. 17.

Thus for instance figs. 13 and 14 must be con-
sidered together. Particularly the sequence shown on the
first line on the top of fig. 13 overlaps with the
35 sequence shown on the first line on the top of fig. 14. In
other words the starting of the reading of the successive

sequences of the env gene as represented in figs. 13-18 involves first reading the first line at the top of fig. 13 then proceeding further with the first line of fig. 14. One then returns to the beginning of the second line of fig. 13, then again further proceed with the reading of the second line of page 14, etc... The same observations then apply to the reading of the paired figs. 15 and 16, and paired figs. 17 and 18, respectively.

The locations of neutralizing epitopes are further apparent in figs. 13-18. reference is more particularly made to the boxed groups of three letters included in the aminoacid sequences of the envelope proteins (reading phase 3) which can be designated generally by the formula N-X-S or N-X-T, wherein X is any other possible aminoacid. Thus the initial protein product of the env gene in a glycoprotein of molecular weight in excess of 91,000. These groups are deemed to generally carry glycosylated groups. These N-X-S and N-X-T groups with attached glycosylated groups form together hydrophylic regions of the protein and are deemed to be located at the periphery of and to be exposed outwardly with respect to the normal conformation of the proteins. Consequently they are considered as being epitopes which can efficiently be brought into play in vaccine compositions.

The invention thus concerns with more particularly peptide sequences included in the env-proteins and excizable therefrom (or having the same aminoacid structure), having sizes not exceeding 200 aminoacids.

Preferred peptides of this invention (referred to hereafter as a, b, c, d, e, f) are deemed to correspond to those encoded by the nucleotide sequences which extend respectively between the following positions :

- a) from about 8085 to about 8200
- b) " " 8280 " " 8310
- 35 c) " " 8390 " " 8440
- d) " " 8465 " " 8620

e) " " 6880 " " 6930
 f) " " 7535 " " 7630

Other hydrophilic peptides in the env open reading frame are identified hereafter. they are defined starting from
 5 aminoacid 1 = lysine (K) coded by the AAA at position 5670-2 in the LAV DNA sequence.

These hydrophilic peptides are
 8-23 aminoacids inclusive

10	63-78	"	"
	82-90	"	"
	97-123	"	"
	127-183	"	"
	197-201	"	"
15	239-294	"	"
	300-327	"	"
	334-381	"	"
	397-424	"	"
	488-500	"	"
20	510-523	"	"
	551-577	"	"
	594-603	"	"
	621-630	"	"
	657-679	"	"
25	719-758	"	"
	780-803	"	"

The invention also relates to any combination of these peptides.

4) The other ORF

30 The invention further concerns DNA sequences which provide open reading frames defined as ORF-Q, ORF-R and as "1", "2", "3", "4", "5", the relative position of which appears more particularly in figs. 2 and 3.

These ORFs have the following locations :

35	ORF-Q	phase 1	start 4478	stop 5086
	ORF-R	" 2	" 8249	" 8886

ing
ing

13

ORF-1	-	1	-	5029	-	5316
ORF-2	-	2	-	5273	-	5515
ORF-3	-	1	-	5383	-	5516
ORF-4	-	2	-	5519	-	5773
5 ORF-5	-	1	-	7986	-	8279

The LTR (long terminal repeats) can be defined as lying between position 8560 and position 180 (and extending over position 9097/1). As a matter of fact the end of the genome is at 9097 and, because of the LTR structure of the retrovirus, links up with the beginning of the sequence :

Hind III

CTCAATAAAGCTTGCTTG



15

9097 1

The invention concerns more particularly all the DNA fragments which have been more specifically referred to hereabove and which correspond to open reading frames. It will be understood that the man skilled in the art will be able to obtain them all, for instance by cleaving an entire DNA corresponding to the complete genome of a LAV species, such as by cleavage by a partial or complete digestion thereof with a suitable restriction enzyme and by the subsequent recovery of the relevant fragments. The different DNAs disclosed in the earlier mentioned British Application can be resorted to also as a source of suitable fragments. The techniques disclosed hereabove for the isolation of the fragments which were then included in the plasmids referred to hereabove and which were then used for the DNA sequencing can be used.

Of course other methods can be used. Some of them have been exemplified in the earlier British Application. reference is for instance made to the following methods.

a) DNA can be transfected into mammalian cells with appropriate selection markers by a variety of techniques, calcium phosphate precipitation, polyethylene

glycol, protoplast-fusion, etc..

b) DNA fragments corresponding to genes can be cloned into expression vectors for *E. coli*, yeast- or mammalian cells and the resultant proteins purified.

5 c) The proviral DNA can be "shot-gunned" (fragmented) into procaryotic expression vectors to generate fusion polypeptides. Recombinant producing antigenically competent fusion proteins can be identified by simply screening the recombinants with antibodies against LAV
10 antigens.

The invention also relates more specifically to cloned probes which can be made starting from any DNA fragment according to this invention, thus to recombinant
15 DNAs containing such fragments, particularly any plasmids amplifiable in procaryotic or eucaryotic cells and carrying said fragments.

Using the cloned DNA fragments as a molecular hybridization probe - either by marking with radionucleotides or with fluorescent reagents - LAV virion RNA may be
20 detected directly in the blood, body fluids and blood products (e.g. of the antihemophylic factors such as factor VIII concentrates) and vaccines, i.e. hepatitis B vaccine. It has already been shown that whole virus can be detected in culture supernatants of LAV producing cells. A
25 suitable method for achieving that detection comprises immobilizing virus onto said a support e.g. nitrocellulose filters, etc., disrupting the virion and hybridizing with labelled (radiolabelled or "cold" fluorescent- or enzyme-labelled) probes. Such an approach has already been
30 developed for Hepatitis B virus in peripheral blood (according to SCOTTO J. et al. Hepatology (1983), 3, 379-384).

Probes according to the invention can also be used for rapid screening of genomic DNA derived from the tissue
35 of patients with LAV related symptoms, to see if the proviral DNA or RNA is present in host tissue and other

tissues.

A method which can be used for such screening comprise the following steps : extraction of DNA from tissue, restriction enzyme cleavage of said DNA, electrophoresis of the fragments and Southern blotting of genomic DNA from tissues, subsequent hybridization with labelled cloned LAV proviral DNA. Hybridization in situ can also be used.

Lymphatic fluids and tissues and other non-lymphatic tissues of humans, primates and other mammalian species can also be screened to see if other evolutionary related retrovirus exist. The methods referred to hereabove can be used, although hybridization and washings would be done under non stringent conditions.

The DNA according to the invention can be used also for achieving the expression of LAV viral antigens for diagnostic purposes.

The invention also relates to the polypeptides themselves which can be expressed by the different DNAs of the inventions, particularly by the ORFs or fragments thereof, in appropriate hosts, particularly procaryotic or eucaryotic hosts, after transformation thereof with a suitable vector previously modified by the corresponding DNAs.

These polypeptides can be used as diagnostic tools, particularly for the detection of antibodies in biological media, particularly in sera or tissues of persons afflicted with pre-AIDS or AIDS, or simply carrying antibodies in the absence of any apparent disorders. Conversely the different peptides according to this invention can be used themselves for the production of antibodies, preferably monoclonal antibodies specific of the different peptides respectively. For the production of hybridomas secreting said monoclonal antibodies conventional production and screening methods are used. These monoclonal antibodies, which themselves are part of

the invention then provide very useful tools for the identification and even determination of relative proportions of the different polypeptides or proteins in biological samples, particularly human samples containing
5 LAV or related viruses.

Thus all of the above peptides can be used in diagnostics as sources of immunogens or antigens free of viral particles, produced using non-permissive systems, and thus of little or no biohazard risk.

10 The invention further relates to the hosts (prokaryotic or eucaryotic cells) which are transformed by the above mentioned recombinants and which are capable of expressing said DNA fragments.

Finally it also relates to vaccine compositions
15 whose active principle is to be constituted by any of the expressed antigens, i.e. whole antigens, fusion polypeptides or oligopeptides in association with a suitable pharmaceutical or physiologically acceptable carrier.

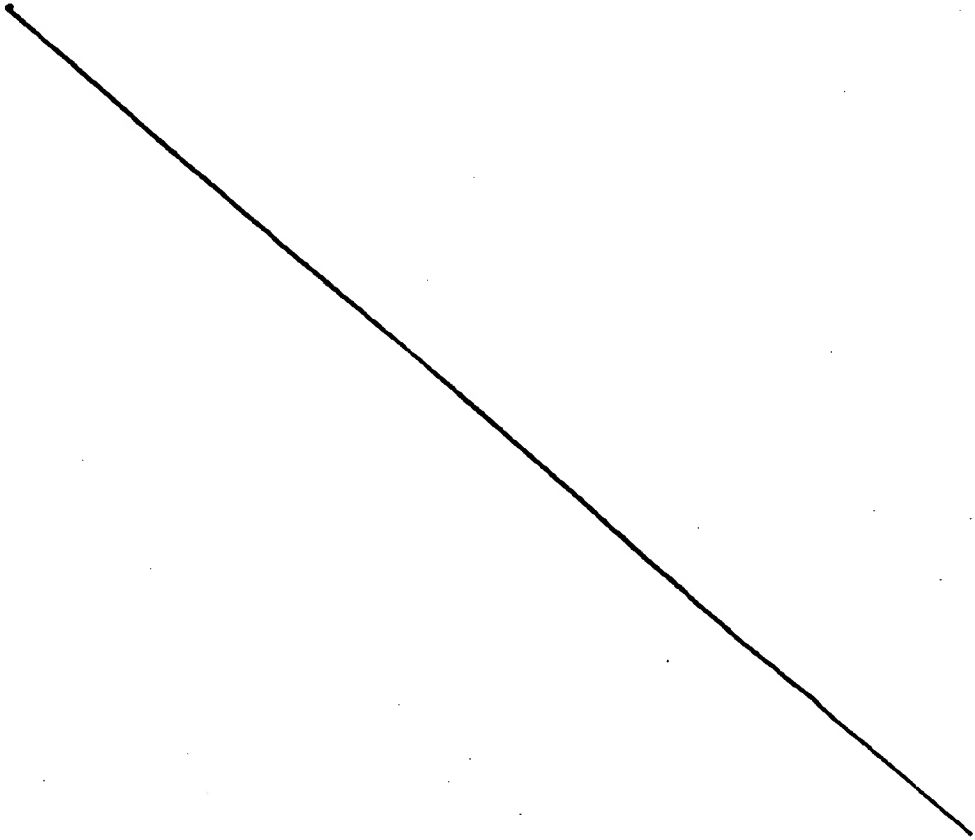
Preferably the active principles to be considered
20 in that field consist of the peptides containing less than 250 aminoacid units, preferably less than 150 as deducible for the complete genomes of LAV, and even more preferably those peptides which contain one or more groups selected from N-X-S and N-X-T as defined above. Preferred peptides
25 for use in the production of vaccinating principles are peptides (a) to (f) as defined above. By way of example having no limitative character, there may be mentioned that suitable dosages of the vaccine compositions are those which enable administration to the host.
30 particularly human host ranging from 10 to 500 micrograms per kg, for instance 50 to 100 micrograms per kg.

For the purpose of clarity figs. 19 to 26 are added. reference may be made thereto in case of difficulties of reading blurred parts of figs. 4 to 12.

Needless to say that figs. 19-25 are merely a reiteration of the whole DNA sequence of the LAV genome.

Finally the invention also concerns vectors for the transformation of eucaryotic cells of human origin, particularly lymphocytes, the polymerases of which are capable of recognizing the LTRs of LAV. Particularly said vectors are characterized by the presence of a LAV LTR therein, said LTR being then active as a promoter enabling the efficient transcription and translation in a suitable host of the above defined, of a DNA insert coding for a determined protein placed under its controls.

Needless to say that the invention extends to all variants of genomes and corresponding DNA fragments (ORFs) having substantially equivalent properties, all of said genomes belonging to retroviruses which can be considered as equivalents of LAV.



CLAIMS

1. A DNA fragment of LAV extending from nucleotide position 236 to nucleotide position 1759.

2. A DNA fragment of LAV extending from nucleotide position 1555 to nucleotide position 5086.

3. A DNA fragment of LAV extending from nucleotide position 5670 to nucleotide position 8132.

4. A vector containing a DNA fragment according to any of claims 1 to 3.

5. Peptide corresponding to any of those encoded by the nucleotide sequences which extend respectively between the following positions :

a) from about 6095 to about 6200

b) " " 6260 " " 6310

c) " " 6390 " " 6440

d) " " 6485 " " 6620

e) " " 6860 " " 6930

f) " " 7535 " " 7630

6. Peptide characterized by a sequence of aminoacids deducible from LAV DNA the terminal aminoacids of which extend between the following positions with respect to the lysine (position 1) coded by the AAA at position 5670-5672 in the LAV DNA.

8-23 aminoacids inclusive

83-78 " "

82-90 " "

97-123 " "

127-183 " "

197-201 " "

238-294 " "

300-327 " "

334-381 " "

397-424 " "

466-500 " "

510-523 " "

551-577 " "

594-603 - -
 621-630 - -
 657-678 - -
 719-758 - -
 780-803 - -

or any combination of these peptides.

7. Peptide corresponding to the aminoacid sequences deducible from LAV DNA and the terminal aminoacids of which are positionned at the positions hereafter counted from the Met at position 1 coded by the ATG sequence at nucleotide positions 260-2 :

12-32 aminoacids inclusive

37-48 - -
 49-79 - -
 88-153 - -
 158-185 - -
 178-188 - -
 200-220 - -
 226-234 - -
 239-264 - -
 268-331 - -
 352-361 - -
 377-390 - -
 398-432 - -
 437-484 - -
 482-498 - -

and combination of said peptides.

8. Diagnostic means containing any of the DNA fragments of any of claims 1 to 3.

9. Diagnostic means containing any of the peptides of any of claims 4 to 6.

10. Vaccine compositions containing any of the peptides according to any of claims 4 to 6 in association with a pharmaceutical vehicle.

End of transmission

Restriction map of the pTZ19 vector. The map shows 10 restriction sites and their positions in base pairs (bp) around a 1000 bp circular plasmid.

Restriction Site	Position (bp)
SalI	0
PstI	116
HindIII	216
BglII	346
KpnI	416
EcoRI	546
KpnI	616
BglII	716
HindIII	816
SacI	916

DFA 2/26

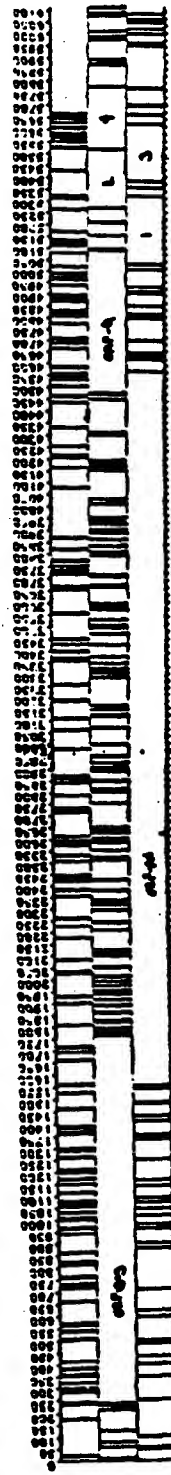
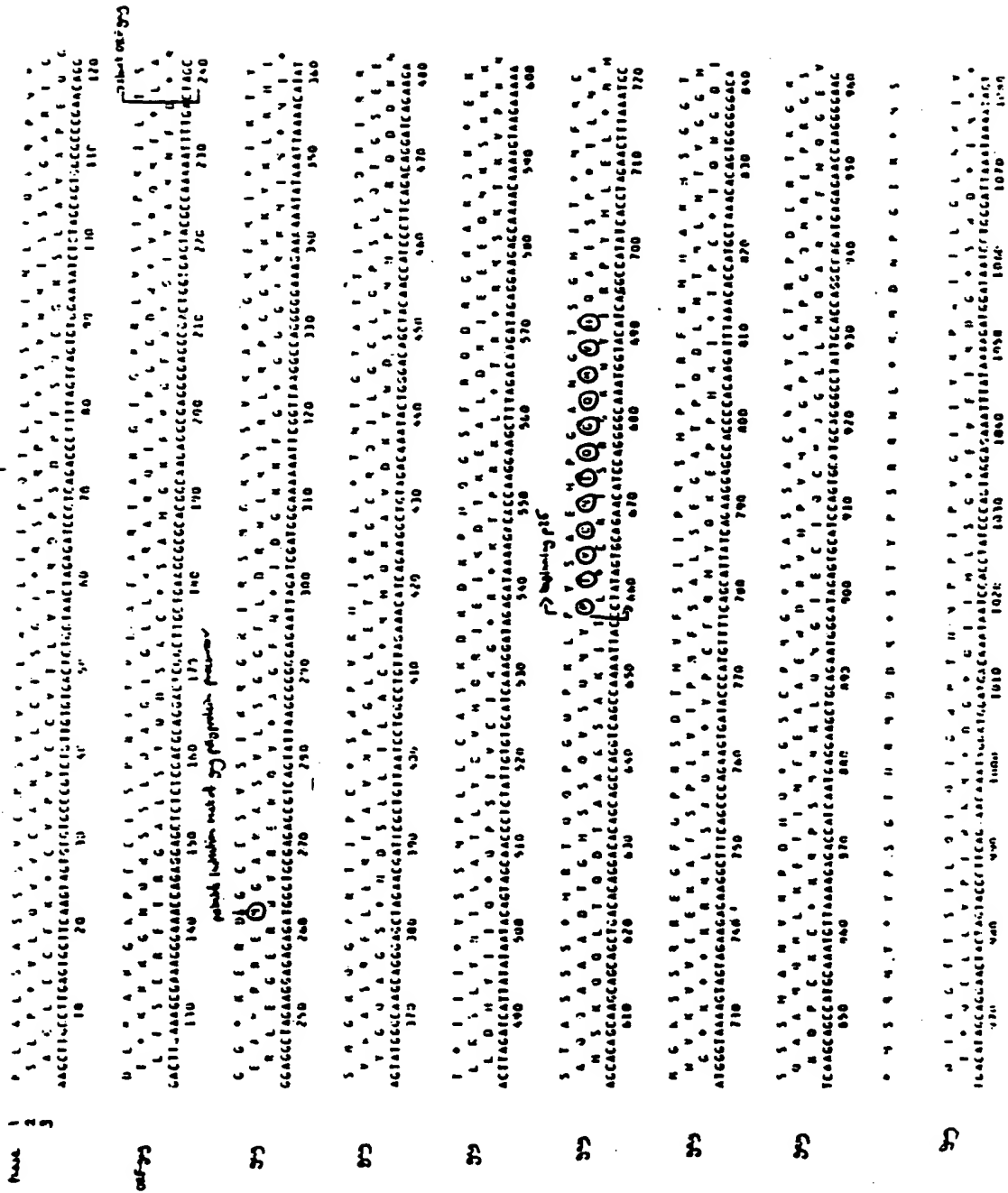


Fig. 2

Year	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	

3
L.V.



16 NOV. 84- 23099

DFA 6126

1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1

Fi. 8

7/26

2.1.2

16 NOV. 34- 29099

D F A

10/26

[illegible]

U. C. 7

16 NOV.84- 29099

DFA 11/26

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2

12/26

0100

16 NOV 87 23099

CAGAGGAGACCAAGAAATGAGGAGTAGATCTAGACTAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTAA
5240 5300 5310 5320 5330 5340 5350

P S L F H N K S L R H L L W O E L A E T A T K T S
O V C F T T K A L G I S Y G R K K R R Q R R R P P
K F V S O Q K P * A S P M A G R S G D S D E D L I
CCAAGTTTGTTCACAACAAAGCCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGACAGCCAGCAGCCTCC
5410 5420 5430 5440 5450 5460 5470

S T C N A T Y T N S N S S I S S S N N N S N S C V
V H V M O P I Q I A I A A L V V A I I I A I V V W
Y * C N L Y K * Q * Q H * * * O * * * O * L C C
AGTACATGTAATGCAACCTATACAAATAGCAATAGCAGCATTAGTAGTAGCAATAATAATAGCAATAGTTGTGTG
5530 5540 5550 5560 5570 5580 5590

U V N * T N R K S R R O W O * E * R R N I S
I D L I D R L I E R A E D S G N E S E G E I S A
* T G * L I D * K E O K T V A M R V K E K Y U
AATAGACAGTTAATTGATAGACTAATAGAAAGAGCAGAGACAGTGGCAATGAGAGTGAAGGAGAAATATCAGC
5650 5660 5670 5680 5690 5700 5710

Y * S V V L O K N C G S O S I M G Y L C G R K Q
I D D L * C Y R K I V G H S L L W G T C V E G S N
L M I C S A T E K L W V T V Y Y G V P V W K E A
TATTGATGATCTGTAGTCTACAGAAAATGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAAC
5720 5730 5740 5750 5760 5770 5780

K Y I F G P H M P V Y P U T P T H K K * Y W * M
G T * C L G H T C L C T H R P O P T R S S I G Y C
V H N V W A T H A C V P T D P N P O E V V L V N
ACGTACATAATGTTTGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTAGTATTGCTAAATGT
5820 5830 5840 5850 5860 5870 5880

C H R I * S V Y G I K A * S H V * N * P H S V L V
A * G Y N O F M G S K P K A M C K I N P T L C * F
H E D I I S L * D O S L K P C V K L T P L C V S L
TGCATGAGGATAAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAATAATTAACCCCACTCTGTGTAGTT
5910 5920 5930 5940 5950 5960 5970

I P I V V A G K * * W R K E R * K T A L S I S A O
Y O * * K G H D D G E R R D K K I L L F O Y O H K
T M S S S G E H N M E K G E I K K N C S F N I S T
ATACCAATAGTACTAGCCGGGAAATGATGATGAGAAAGCAGAGATAAAAACTGCTCTTTCAATATCAGCAGAAC
6130 6140 6150 6160 6170 6180 6190

L I * Y O * I M I L P A I R * U V V T P O S L H R
* Y N T H R * * Y Y O L Y V D K L * H L S H Y T G
U I I P I D V D T T S Y T L T S C N T S V I T O
TTGATATAATACCAATAGATAATGATACTACCAGCTATACGTTGACAAATTTGTAACAGCTCAGTCATTACAGGL
6250 6260 6270 6280 6290 6300 6310

P L V L R F * N V I K R S M E O D H V O M S A

16 NOV 84-529099

S O P K T A C T T C Y C K K C C F H C
V S L K L L V P L A I V K S V A F I A
CAGGAAGTCAGCCTAAAACTGCTTGTACCACTTGTATTGTAAAAAGTGTTCCTTTTCATTG
5350 5360 5370 5380 5390 5400

A T K T S S P Q S D S S S F S I K A V S
U R R R P P Q G S G T H Q V S L S K O * V
S D E D L L K A V R L I K F L Y Q S S K *
AGCGACGAAGACCTCCTCAAGGCAGTCAGACTCATCAAGTTTCTCTATCAAAGCAGTAAGT
5470 5480 5490 5500 5510 5520

S N S C V V H S N H R I * E N I K T K K
I A I V V W S I V I I E Y R K I L R O R K
* Q * L C G P * * S * N I G K Y * D K E K
TAGCAATAGTTGTGTGGTCCATAGTAATCATAGAATATAGGAAAATATTAAGACAAAGAAA
5590 5600 5610 5620 5630 5640

R R N I S T C G D G G G N G A P C S L G
K E I S A L V E H G V E H G H A P W D
K E K Y Q H L W R W G W K W G T H L L G I
AAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAAATGGGGCACCATGCTCCTTGGGA
5710 5720 5730 5740 5750 5760

C G F K Q P P L Y F V H Q M L K H M I Q
V E G S N H H S I L C I R C * S I * Y R
V W K E A T T T L F C A S D A K A Y D T E
TGTGGAAGGAAGCAACCACCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAG
5830 5840 5850 5860 5870 5880

* Y * M * Q K I L T C G K M T W * N R
S I G K C D R K F * H V E K * H G R T D
V V L Y N V T E N F N M W K N O H V E O H
TAGTATTGGTAAATGTGACAGAAAATTTTAACATGTGGAAAAATGACATGGTAGAACAGA
5950 5960 5970 5980 5990 6000

H S V L V * S A L I W G * L L I P I V V
T L C * F K V H * F G E C Y * Y O * *
P L C V S L K C T D L G N A T N T N S S N
CACTCTGTGTTAGTTTAAAGTGCAGTGTATTTGGGGATGCTACTAATACCAATAGTAGTA
6070 6080 6090 6100 6110 6120

S I S A Q A * E V R C P K N M H F F I N
O Y O H K H K R * G A E R I C I F L * T
F N I S T S I R G K V G K E Y A F F Y K L
TCAATATCAGCACAAGCATAAGAGGTAAGGTCCAGAAAGAATATGCATTTTTTTATAAAC
6190 6200 6210 6220 6230 6240

S L H R P V O R Y P L S Q F P Y I I V
S H Y T G L S K G I L * A N S H T L L C
S V I T Q A C P K V S F E P I P I H Y C A
CAGTCATTACACAGGCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTG
6310 6320 6330 6340 6350 6360

V Q M S A Q Y N V H * F L G O * Y Q L N

16 NOV 84 29039

DEF 15

12/26

P G A F C D S A * * * * * V J A H R T M Y K C *
P A G F A I L K C N [N K T] F [N G S] G P C T [V V V]
CCCCGGCTGGTTTTCGATTCTAAAATGTAATATAAGACGTTCAATGGAAACAGGACCATGTACAAAATCTCAG
6370 6380 6390 6400 6410 6420 6430

C C * M A V * Q K K R * * L D L P I S O T N L K I
A V E W O S S R R R G S N * I C O F H K O C * N
L L [N G S] L A E E E V V I R S A [N F T] D N A K Y
TCTGTTCAATGGCAGTCTAGCAGAAGAAGGAGGTAGTAATAGATCTCCCAATTTACAGACAATGCTAAAACG
6490 6500 6510 6520 6530 6540 6550

P T T I G E K V S V S R G D O G E H L L O * E K *
O C O Y K K K Y P Y P E G T R E S I C Y N P K N
N [N H Y] R K S I R I O R G P G R A F V T I G K I
CCAACAACAATACAAGAAAAAGTATCCGTATCCAGAGGGGACCAGGGAGAGCATTGTGTACAATAGCAAAAAATA
6610 6620 6630 6640 6650 6660 6670

M P L * N R * L A N * E N N L E I I K O * S L S G
C H F K T O S * O I K R T I W K * * N N N L * A
[A T] L K O I A S K L R E O F G N [N K T] I I F K O
ATGCCACTTTAAAACAGATAGCTAGCAATTAAGAGAACAATTTGGAATAATAAAACAATAATCTTTAAGCAA
6730 6740 6750 6760 6770 6780 6790

I G N F S T V I O H N C L I V L G L I V L G V L K
H G I F L L * F N T T V * * Y L V * * Y L E Y *
G E F F Y C [N S T] Q L F [N S T] W F [N S T] W S T E
GAGGGGAATTTTCTACTGTAATTCACACAAGTGTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTGAAC
6850 6860 6870 6880 6890 6900 6910

E * N N L * T C G R K * E K O C M P L P S A D K L
N K T I Y K H V A G S R K S N V C P S H Q R T H *
I K O F I N M W O E V G K A M Y A P P I S G O I
GAATAAAACAATTTATAAACATGTGGCAGGAAGTAGAAAAAGCAATGTATGCCCTCCCATCAGCGGACAAATTA
6970 6980 6990 7000 7010 7020 7030

V I T T H G P R S S D L E E E I * G T I G E V N Y
* * O O W V R D L O T W R R R Y E G O L E K * I I
N N N [N G S] E I F R P G G G D H R O N W R S E L
GTAATAACAACAATGGGTCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTTGAGAAGTGAATTA
7090 7100 7110 7120 7130 7140 7150

P R O R E E W C R E K K E O W E * E L C S L G S W
O G K E K S G A E R K K S S G N R S F V P M V L
K A K R R V V O R E K R A V G I G A L F L G F L
CCAAGGCAAAAGAGAAGTGGTGCAGACAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCTTGGCTTTCG
7210 7220 7230 7240 7250 7260 7270

Y R P D N Y C L V * C S S R T I C * G L L R R N S
T G O T I I V H Y S A A A E O F A E G Y * G A T
O A R O L L S G I V O O O N N L L R A I E A O O
TACAGGCCAGACAATTTGCTGCTATAGTGCACGACGAGAACAATTTGCTGAGGGCTATTGAGGCCCAACAGC
7330 7340 7350 7360 7370 7380 7390

E S A L M K D T * R I N S S W G F G V A L E N S F

16 NOV. 84 23299

DFA

16/26

H T M Y K C H S T M Y T M Y A S S I N S
G P C T N V S T V O C T H G I R P V V S T U L
AACAGGACCATGTACAAATGTCAGGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAAC
6420 6430 6440 6450 6460 6470 6480

I S O T M L K P * * Y S * T N L * K L I V U O
O F H R O C * N H N S T A E P I C K N * L Y K T
N F T D N A K T I I V O L N O S V E I N C T R P
CAATTTACAGACAATGCTAAAACCATAATAGTACAGCTGAACCAATCTGTAGAAATTAATTGTACAAGAC
6540 6550 6560 6570 6580 6590 6600

F H L L O * E K * E I * O K H I V T L V F O N G
S I C Y N P K N R K Y E T S T L * H * S K M E
A F V T I G K I G N M R O A H C P I S R A K W N
AGCATTGTGTACAATAGGAAAAATAGGAAATATGAGACAAGCACATTGTAACATTAGTAGAGCAAAATGCA
6660 6670 6680 6690 6700 6710 6720

I I K O * S L S N P O E G T O K L * R T V L I V
* * N * V L * A I L R R G P R N C N A O F * L W
N K T I I F K O S S G G D P E I V T H S F N C G
TAATAAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAACGCACAGTTTAAATTGTC
6780 6790 6800 6810 6820 6830 6840

G L I V L G V L K G Q I T L K E V T O S M S H A
V * * Y L E Y * R V K * M * R K * H N H T P M O
F N S T W S T E G S N N T E G S O T I T L P C R
GTTTAATAGTACTTGGAGTACTGAAGGGTCAAAATAACACTGAAGGAAGTGACACAATCACACTCCCATGCA
6900 6910 6920 6930 6940 6950 6960

M P L P S A D K L D V H O I L O G C Y * O E M V
C P S H O R T N * M F I K Y Y R A A I N K R W W
A P P I S G O I R C S S N I T G L L L T R O G G
TGCCCCCTCCCATCAGCGGACAAATTAGATGTTTCATCAAAATATTACAGGGCTGCTATTAACAAGAGATGGTG
7020 7030 7040 7050 7060 7070 7080

* G T I G E V N Y I N I K * * K L N H * E * H P
E G O L E K * I I * I * S S K N * T I R S S T H
R O N W R S E L Y K Y K V V K I E P L G V A P T
GAGGCACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCA
7140 7150 7160 7170 7180 7190 7200

* E L C S L G S W E O O E A L W A H G O * R * R
R S F V P W V L G S S R K H Y G R T V N O A O G
G A L F L G F L G A A G S T M G A R S M T L T V
AGGAGCTTTGTTCTTGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCACGGTCAATGACGCTGACGG
7260 7270 7280 7290 7300 7310 7320

* G L L R R N S I C C N S O S G A S S S S R O
A E G Y * G A T A S V A T H S L G H O A A P C K
L R A I E A O O H L L O L T V W G I K U L O A R
CTGAGCGCTATTGAGCGGCAACAGCATCTGTTCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCA
7380 7390 7400 7410 7420 7430 7440

G V A L E N S F A P L L C L G * L V G V I N L

16 NOV. 84-29899

V P G C G K I P K G S T A P G D L G L L A K T H
I L A V E R Y L K D O G L L G I W G C S G K L I
GAATCCTGGCTGTCGAAGATACCTAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTGGAAAACCTCATI
7450 7460 7470 7480 7490 7500 7510

M N R F G I T * P G W S G T E K L T I T O A * Y I
G T O L E * H D L D G V G D R N * O L H K L N T
E D I W N N T W M E W D R E I N N Y T S L I H
TGAACAGATTGGGAATAACATGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACAT
7570 7580 7590 7600 7610 7620 7630

N Y * H * I N G O V C G I G L T * G I G C G I * K
I I G I R * M G K F V E L V * W N K L A V V Y K
L L E L D K W A S L W N W F N I T N W L W Y I K
AATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGCTTAACATAACAAATTGGCTGTGGTATATAAAA
7690 7700 7710 7720 7730 7740 7750

L L Y F L * * I E L G R D I H H Y R F R P T S Q P
C C T F Y S E * S * A G I F T I I V S D P P P N
A V L S I V N R V R O G Y S P L S F O T H L P T
TTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATCACCATTATCGTTTCAGACCCACCTCCCAACC
7810 7820 7830 7840 7850 7860 7870

R E T E T D P F D * * T D P * H L S G T I C G A L
E R U P Q I H S I S E R I L S T Y L G R S A E P
R D R D R S I R L V N G S L A L I W D D L R S L
AGAGAGACAGACAGATCCATTGATTAGTGAACGATCTTAGCATTATCTGGGACGATCTGCGGAGCCTT
7930 7940 7950 7960 7970 7980 7990

T R I V E L L G K R G W E A L K Y W W N L L O Y N
R G L W N F W D A G C G K P S N I G G I S Y S I
E D C G T S G T G G V G S P O I L V E S P T V L
ACGAGGATTGTGGAACCTTCTGGGACGACGGGGTGGGAAGCCCTCAAATATTGCTGGAATCTCTACAGTATTC
8050 8060 8070 8080 8090 8100 8110

A I A V A E G T O R V I E V V O G A C R A I R H I
P * J * L R G Q I G L * K * Y K E L V E L F A T
H S S S * G D R * G Y R S S T R S L * S Y S P H
GCCATAGCAGTAGCTGAGGGGACAGATAGGCTTATAGAAGTAGTACAAGGAGCTGTAGAGCTATTGCCACAT
8170 8180 8190 8200 8210 8220 8230

G W O V V K K * C G W H A Y C K G K V E T S * A S
C G K W S K S S V V G W P T V R E R M R R A E P
Y A S G O K V W L D G L L * G K E * D E L S O
GGTGCGCAAGTGGTCAAAAAGTAGTGTGGTTGGATGGCCTACTGTAAGGCAAGAATGAGACGAGCTGAGCCAG
8290 8300 8310 8320 8330 8340 8350

S N H K * O Y S S Y O C C L C L A R S T R G G C
A I T S S N T A A T N A A C A W L F A O E E E
U S O V A I U D L P M L L V P G * K H K R R R S
ACCAATCACAAGTAGCAATACAGCAGCTACCAATGCTGCTTGTGCTTGGCTTGAAGGACAGAGGAGGAGGAGG
8410 8420 8430 8440 8450 8460 8470

U G S C R S * P L F K R K G C T G

using 15/15

16 NOV 84 29099

13/18

DFA

A K T H L H H C C A L E C * L E * * I S
K L I C T T A V P W V A S W S W S L
CTGGAAACTCATTTGCACCACTGCTGCGCTTGGATGCTAGTTGGAGTAATAAATCTC
7510 7520 7530 7540 7550 7560

O A * Y I P * L K N R K T S K K R M N K
K L N T F L N * R I A K P A R K E * T R
S L I H S L I E E S O V O O E K N E O E
CAAGCTTAATACATTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAG
7630 7640 7650 7660 7670 7680

C G I * K Y S * * * E A W * V * E * F
V V Y K N I H N D S R R L G R F K N S F
W Y I K I F I M I V G G L V G L / R / I V F
TGTGGTATATAAAATATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTT
7750 7760 7770 7780 7790 7800

P T S Q P R G D P T G P K E * K K K V E
P P P N P E G T R O A R R N R R R R W R
T H L P T P R G P D R P E G I E E E G G E
CCCACCTCCCAACCCGAGGGGACCCGACAGGCCCAAGGAATAGAAGAAGAAGGTGGAG
7870 7880 7890 7900 7910 7920

I C G A L C L F S Y H R L R D L L L I V
S A E P C A S S A T T A * E T Y S * L *
L R S L V P L O L P P L E R L T L D C N
ATCTGCGGAGCCTTGTGCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTGTA
7990 8000 8010 8020 8030 8040

L L O Y H S O E L K N S A V S L L N A T
S Y S I G V R N * R I V L L A C S M P O
P T V L E S G T K E * C C * L A O C H S
TCCTACAGTATTGGAGTCAGGAATAAGAATAGTGCTGTTAGCTTGCTCAATGCCACA
8110 8120 8130 8140 8150 8160

A I R H I P R R I R O G L E R I L L * D
L F A T Y L E E * D R A W K G F C Y K M
Y S P H T * K N K T G L G K D F A I R W
CTATTGCCCACATACCTAGAAGAATAAGACAGGCTTGGAAAGGATTTTGTATAAGAT
8230 8240 8250 8260 8270 8280

T S * A S S R A G G S S I S R P G K T W
R A E P A A D G V G A A S R D L E K H G
E L S O O O * G W E O H L E T W K N M E
CGAGCTGAGCCAGCAGCAGATGGGGTGGGAGCAGCATCTCGAGACCTGGAAAAACATGG
8350 8360 8370 8380 8390 8400

R G C G G G F S S H T S G T F K T N D L
E E E E V G F P V T P C V P L R P M T Y
R R P R N Y F J S H L R Y L * D O * L T
AGCAGCAGCAGCAGGCGGTTTCCAGTCAGCCTCAGCTACCTTTAAGACCAATCACTTA
8450 8460 8470 8480 8490 8500 8510 8520

3 L P T B 1/4
15/15 B 1/4
S V G L P H T R L L

16 NOV. 64- 28095-

19/26

D F A

Fig 19

10	20	30	40	50	60
AAGCTTGCCT	TGAGTGCCTC	AAGTAGTGTC	TGCCCCGTCTG	TTGTGTGACT	CTCGTAACTA
70	80	90	100	110	120
GAGATCCCTC	AGACCCCTTT	AGTCAGTGTC	GAAAACTCT	AGCAGTGGCG	CCCCAACAGG
130	140	150	160	170	180
GACTTGAAAG	CGAAAGGGAA	ACCAGAGGAG	CTCTCTCGAC	GCAGGACTCG	GCTTGCTGAA
190	200	210	220	230	240
CGCGCCACGG	CAAGAGGCGA	GGGGAGGCGA	CTGGTGAGTA	CGCCAAAAAT	TTTGACTAGC
250	260	270	280	290	300
GGAGGCTAGA	AGGAGAGAGA	TGGGTGCGAG	AGCGTCAGTA	TTAAGCGGGG	GAGAATTAGA
310	320	330	340	350	360
TGGATCGGAA	AAAATTCGGT	TAAGGCCAGG	GGGAAAGAAA	AAATATAAAT	TAAACATAT
370	380	390	400	410	420
AGTATGGGCA	AGCAGGGAGC	TAGAACGATT	CGCTGTTAAT	CCTGGCCTGT	TAGAAACATC
430	440	450	460	470	480
AGAAGGCTGT	AGACAAATAC	TGGGACAGCT	ACAACCATCC	CTTCAGACAG	GATCAGAAGA
490	500	510	520	530	540
ACTTAGATCA	TTATATAATA	CAGTAGCAAC	CCTCTATTGT	GTGCATCAAA	GGATAGAGAT
550	560	570	580	590	600
AAAAGACACC	AAGGAAGCTT	TAGACAAGAT	AGAGGAAGAG	CAAAACAAAA	GTAAGAAAAA
610	620	630	640	650	660
AGCACAGCAA	GCAGCAGCTG	ACACAGGACA	CAGCAGCCAG	GTGAGCCAAA	ATTACCCTAT
670	680	690	700	710	720
AGTGCAGAAC	ATCCAGGGGC	AAATGGTACA	TCAGGGCATA	TCACCTAGAA	CTTTAAATGC
730	740	750	760	770	780
ATGGGTAAAA	GTAGTACAAG	AGAAGGCTTT	CAGCCCAGAA	GTGATACCCA	TGTTTTCAGC
790	800	810	820	830	840
ATTATCAGAA	GGAGCCACCC	CACAAGATTT	AAACACCATG	CTAAACACAG	TGGCGGGACA
850	860	870	880	890	900
TCAAGCAGCC	ATGCAAATGT	TAAAAGAGAC	CATCAATGAG	GAAGCTGCAG	AATGGGATAG
910	920	930	940	950	960
AGTGCATCCA	GTGCATGCAG	GGCCTATTGC	ACCAGGCCAG	ATGAGAGAAC	CAAGGGGAAG
970	980	990	1000	1010	1020
TGACATAGCA	GGAACACTA	GTACCCTTCA	GGAACAAATA	GGATGGATGA	CAAATAATCC
1030	1040	1050	1060	1070	1080
ACCTATCCCA	GTAGGAGAAA	TTTATAAAAG	ATGGATAATC	CTGGGATTAA	ATAAAATAGT
1090	1100	1110	1120	1130	1140

056/2

Fig 20

AAGAATCTAT	AGCCCTACCA	GCATTCTGGA	CATAAGACAA	GGACCAAAAG	AACCCCTTAC
1150	1160	1170	1180	1190	1200
AGACTATGTA	GACCGGTCT	ATAAACTCT	AAGAGCCGAG	CAAGCTTCAC	AGGAGGTAAA
1210	1220	1230	1240	1250	1260
AAATTGGCATG	ACAGAAACCT	TGTTGGTCCA	AAATGCCAAC	CCAGATTGTA	AGACTATTTT
1270	1280	1290	1300	1310	1320
AAAAGCATTG	CGACCAGCAG	CTACACTAGA	AGAAATCATG	ACAGCATGTC	AGGGAGTGGG
1330	1340	1350	1360	1370	1380
AGGACCCGGC	CATAAGGCAA	GAGTTTGGC	TGAAGCAATG	AGCCAAGTAA	CAAATTCACC
1390	1400	1410	1420	1430	1440
TACCATAATC	ATGCAAAGAG	GCAATTTTAG	GAACCAAAGA	AAGATTGTTA	AGTGTTCCTA
1450	1460	1470	1480	1490	1500
TTGTGGCAAA	GAAGGGCACA	TAGCCAGAAA	TTGCAGGGCC	CCTAGGAAAA	AGGGCTGTTG
1510	1520	1530	1540	1550	1560
GAAATGTGGA	AAGGAAGGAC	ACCAAATGAA	AGATTGTACT	GAGAGACAGG	CTAATTTTTT
1570	1580	1590	1600	1610	1620
AGGGAAGATC	TGGCCTTCCT	ACAAGGGAAG	GCCAGGGAAT	TTTCTTCAGA	GCAGACCAGA
1630	1640	1650	1660	1670	1680
GCCAACAGCC	CCACCAGAAG	AGAGCTTCAG	GTCTGGGGTA	GAGACAACAA	CTCCCTCTCA
1690	1700	1710	1720	1730	1740
GAAGCAGGAG	CCGATAGACA	AGGAAGTGTG	TCCTTTAACT	TCCCTCAGAT	CACTCTTTGG
1750	1760	1770	1780	1790	1800
CAACGACCCC	TCGTCACAA	AAAGATAGGG	GGGCAACTAA	AGGAAGCTCT	ATTAGATACA
1810	1820	1830	1840	1850	1860
GGAGCAGATG	ATACAGTATT	AGAAGAAATG	AGTTTGCCAG	GAAGATGGAA	ACCAAAAAATG
1870	1880	1890	1900	1910	1920
ATAGGGGGAA	TTGGAGGTTT	TATCAAAGTA	AGACAGTATG	ATCAGATACT	CATAGAAATC
1930	1940	1950	1960	1970	1980
TGTGGACATA	AAGCTATAGG	TACAGTATTA	GTAGGACCTA	CACCTGTCAA	CATAATTGGA
1990	2000	2010	2020	2030	2040
AGAAATCTGT	TGACTCAGAT	TGGTTGCACT	TTAAATTTTC	CCATTAGTCC	TATTGAAACT
2050	2060	2070	2080	2090	2100
GTACCAGTAA	AATTAAAGCC	AGGAATGGAT	GGCCCAAAAG	TTAAACAATG	GCCATTGACA
2110	2120	2130	2140	2150	2160
GAAGAAAAAA	TAAAAGCATT	AGTAGAAATT	TGTACAGAAA	TGCAAAAGGA	AGGGAAAAAT
2170	2180	2190	2200	2210	2220
TCAAAAAATTG	GGCCTGAAAA	TCCATACAAT	ACTCCAGTAT	TTGCCATAAA	GAaaaaaacac
2230	2240	2250	2260	2270	2280
AGTACTAAAT	GCAGAAAAAT	AGTAGATTTT	AGAGAACTTA	ATAAGAGAAC	TCAAGACTTC
2290	2300	2310	2320	2330	2340
TGGGAAGTTC	AATTAGGAAT	ACCACATGCC	GCAGGGTTAA	AAAAGAAAAA	ATCAGTAACA

16 NOV. 84- 29099

D F A

2410 2420 2430 2440 2450 2460
 ACTGCATT TA CCATACCTAG TATAACAAT GAGACACCAG GGATTAGATA TCAGTACAAT
 2470 2480 2490 2500 2510 2520
 GTGCTTCCAC AGGGATGGAA AGGATCACCA GCAATATTCC AAAGTAGCAT GACAAAAATC
 2530 2540 2550 2560 2570 2580
 TTAGAGCCTT TTAGAAAAACA AAATCCAGAC ATAGTTATCT ATCAATACAT GGATGATTTG
 2590 2600 2610 2620 2630 2640
 TATGTAGGAT CTGACTTAGA AATAGGGCAG CATAGAACAA AAATAGAGGA GCTGAGACAA
 2650 2660 2670 2680 2690 2700
 CATCTGTTGA GGTGGGGACT TACCACACCA GACAAAAAAC ATCAGAAAGA ACCTCCATTG
 2710 2720 2730 2740 2750 2760
 CTTTGGATGG GTTATGAACT CCATCCTGAT AAATGGACAG TACAGCCTAT AGTGCTGCCA
 2770 2780 2790 2800 2810 2820
 GAAAAAGACA GCTGGACTGT CAATGACATA CAGAAGTTAG TGGGAAAATT GAATTGGGCA
 2830 2840 2850 2860 2870 2880
 AGTCAGATTT ACCCAGGGAT TAAAGTAAGG CAATTATGTA AACTCCTTAG AGGAACCAAA
 2890 2900 2910 2920 2930 2940
 GCACTAACAG AAGTAATACC ACTAACAGAA GAAGCAGAGC TAGAACTGCC AGAAAACAGA
 2950 2960 2970 2980 2990 3000
 GAGATTCTAA AAGAACCAGT ACATGGAGTG TATTATGACC CATCAAAAGA CTTAATAGCA
 3010 3020 3030 3040 3050 3060
 GAAATACAGA AGCAGGGGCA AGGCCAATGG ACATATCAAA TTTATCAAGA GCCATTTAAA
 3070 3080 3090 3100 3110 3120
 AATCTGAAAA CAGGAAAAATA TGCAAGAAGC AGGGGTGCCC AACTAATGA TGTAACAAA
 3130 3140 3150 3160 3170 3180
 TTAACAGAGG CAGTGCAAAA AATAACCACA GAAAGCATAG TAATATGGGG AAAGACTCCT
 3190 3200 3210 3220 3230 3240
 AAATTTAAAC TACCCATACA AAAGGAAACA TGGGAAACAT GGTGGACAGA GTATTGGCAA
 3250 3260 3270 3280 3290 3300
 GCCACCTGGA TTCCTGAGTG GGAGTTTGTC AATACCCCTC CTTTAGTGAA ATTATGGTAC
 3310 3320 3330 3340 3350 3360
 CAGTTAGAGA AAGAACCCTAT AGTAGGAGCA GAAACGTTCT ATGTAGATGG GGCAGCTAGC
 3370 3380 3390 3400 3410 3420
 AGGGAGACTA AATTAGGAAA AGCAGGATAT GTTACTAATA GAGGAAGACA AAAAGTTGTC
 3430 3440 3450 3460 3470 3480
 ACCCTAACTG ACACAACAAA TCAGAAGACT GAGTTACAAG CAATTCATCT AGCTTTGCAG
 3490 3500 3510 3520 3530 3540
 GATTCGGGAT TAGAAGTAAA TATAGTAACA GACTCACAAT ATGCATTAGG AATCATTCAA
 3550 3560 3570 3580 3590 3600
 GCACAACCAG ATAAAAGTGA ATCAGAGTTA GTCAATCAAA TAATAGAGCA CTTAATAAAA
 3610 3620 3630 3640 3650 3660

221

Fig 22

15 NOV 84 - 29 099

3670	3680	D F A	3690	3700	3710	3720
GTAGATAAAT	TAGTCAGTGC	TGGAATCAGG	AAAGTACTAT	TTTGTAGATGG	AATAGAT	
3730	3740	3750	3760	3770	3780	
GCCCAAGATG	AACATGAGAA	ATATCACAGT	AATTGGAGAG	CAATGGCTAG	TGATTTTAAC	
3790	3800	3810	3820	3830	3840	
CTGCCACCTG	TAGTAGCAAA	AGAAATAGTA	GCCAGCTGTC	ATAAATGTCA	GCTAAAACGA	
3850	3860	3870	3880	3890	3900	
GAAGCCATGC	ATGGACAAGT	AGACTGTAGT	CCAGGAATAT	GGCAACTAGA	TTGTACACAT	
3910	3920	3930	3940	3950	3960	
TTAGAAGCAA	AAGTTATCCT	GGTAGCAGTT	CATGTAGCCA	GTGGATATAT	AGAAGCAGAA	
3970	3980	3990	4000	4010	4020	
GTTATTCCAG	CAGAAACAGG	GCAGGAAACA	GCATACTTTC	TTTTAAAATT	AGCAGGAAGA	
4030	4040	4050	4060	4070	4080	
TGGCCAGTAA	AAACAATACA	TACAGACAAT	GGCAGCAATT	TCACCAGTAC	TACGGTTAAG	
4090	4100	4110	4120	4130	4140	
GCCGCTGTT	GGTGGCGGG	AATCAAGCAG	GAATTTGGAA	TTCCCTACAA	TCCCCAAAGT	
4150	4160	4170	4180	4190	4200	
CAAGGAGTAG	TAGAATCTAT	GAATAAGAA	TTAAAGAAAA	TTATAGGCCA	GGTAAGAGAT	
4210	4220	4230	4240	4250	4260	
CAGGCTGAAC	ATCTTAAGAC	AGCAGTACAA	ATGGCAGTAT	TCATCCACAA	TTTTAAAAGA	
4270	4280	4290	4300	4310	4320	
AAAGGGGGGA	TTGGGGGGTA	CAGTGCAGGG	GAAAGAATAG	TAGACATAAT	AGCAACAGAC	
4330	4340	4350	4360	4370	4380	
ATACAAACTA	AAGAATTACA	AAAACAAATT	ACAAAAATTC	AAAATTTTCG	GGTTTATTAC	
4390	4400	4410	4420	4430	4440	
AGGCACAGCA	GAGATCCACT	TTGGAAAGGA	CCAGCAAAGC	TCCTCTGGAA	AGGTGAAGGG	
4450	4460	4470	4480	4490	4500	
GCAGTAGTAA	TACAAGATAA	TAGTGACATA	AAAGTAGTGC	CAAGAAGAAA	AGCAAAGATC	
4510	4520	4530	4540	4550	4560	
ATTAGGGATT	ATGGAAAACA	GATGGCAGGT	GATGATTGTC	TGGCAAGTAG	ACAGGATGAG	
4570	4580	4590	4600	4610	4620	
GATTAGAACA	TGGAAAAGTT	TAGTAAAACA	CCATATGTAT	GTTTCAGGGA	AAGCTAGGGG	
4630	4640	4650	4660	4670	4680	
ATGGTTTTAT	AGACATCACT	ATGAAAGCCC	TCATCCAAGA	ATAAGTTCAG	AAGTACACAT	
4690	4700	4710	4720	4730	4740	
CCCACTAGGG	GATGCTAGAT	TGGTAATAAC	AACATATTGG	GGTCTGCATA	CAGGAGAAAG	
4750	4760	4770	4780	4790	4800	
AGACTGGCAT	CTGGGTCAGG	GAGTCTCCAT	AGAATGGAGG	AAAAAGAGAT	ATAGCACACA	
4810	4820	4830	4840	4850	4860	
AGTAGACCT	GAAGTAGCAG	ACCAACTAAT	TCATCTGTAT	TACTTTGACT	GTTTTTCAGA	
4870	4880	4890	4900	4910	4920	

16 NOV 84 - 29099

4930 4940 D F A4950 4960 4970 4980
 AGGACATAAC AAGGTAGGAT CTGTACAATA CTGGGCACTA GCAGCATTAA TAACACCAAA
 4990 5000 5010 5020 5030 5040
 AAAGATAAAG CCACCTTTGC CTAGTGTTAC GAAACTGACA GAGGATAGAT GGAACAAGCC
 5050 5060 5070 5080 5090 5100
 CCAGAAGACC AAGGGCCACA GAGGGAGCCA CACAATGAAT GGACACTAGA GCTTTTAGAG
 5110 5120 5130 5140 5150 5160
 GAGCTTAAGA ATGAAGCTGT TAGACATTTT CCTAGGATTT GGCTCCATGG CTTAGGGCAA
 5170 5180 5190 5200 5210 5220
 CATATCTATG AAACCTTATGG GGATACTTGG GCAGGAGTGG AAGCCATAAT AAGAATTCTG
 5230 5240 5250 5260 5270 5280
 CAACAACCTGC TGTATTATCCA TTTCAGAAAT GGGTGTGGAC ATAGCAGAAT AGGCGTTACT
 5290 5300 5310 5320 5330 5340
 CAACAGAGGA CAGCAAGAAA TGGAGCCAGT AGATCCTAGA CTAGAGCCCT GGAAGCATCC
 5350 5360 5370 5380 5390 5400
 AGGAAGTCAG CCTAAAACCTG CTTGTACCAC TTGCTATTGT AAAAAGTGT GCTTTCATTG
 5410 5420 5430 5440 5450 5460
 CCAAGTTTGT TTCACAACAA AAGCCTTAGG CATCTCCTAT GGCAGGAAGA AGCGGAGACA
 5470 5480 5490 5500 5510 5520
 GCGACGAAGA CCTCCTCAAG GCAGTCAGAC TCATCAAGTT TCTCTATCAA AGCAGTAAGT
 5530 5540 5550 5560 5570 5580
 AGTACATGTA ATGCAACCTA TACAAATAGC AATAGCAGCA TTAGTAGTAG CAATAATAAT
 5590 5600 5610 5620 5630 5640
 AGCAATAGTT GTGTGGTCCA TAGTAATCAT AGAATATAGG AAAATATTAA GACAAAGAAA
 5650 5660 5670 5680 5690 5700
 AATAGACAGG TTAATTGATA GACTAATAGA AAGAGCAGAA GACAGTGGCA ATGAGAGTCA
 5710 5720 5730 5740 5750 5760
 AGGAGAAATA TCAGCACTTG TGGAGATGGG GGTGGAAATG GGGCACCATG CTCCTTGGGA
 5770 5780 5790 5800 5810 5820
 TATTGATGAT CTGTAGTGCT ACAGAAAAAT TGTGGGTCAC AGTCTATTAT GGGGTACCTG
 5830 5840 5850 5860 5870 5880
 TGTGGAAGGA AGCAACCACC ACTCTATTTT GTGCATCAGA TGCTAAAGCA TATGATACAG
 5890 5900 5910 5920 5930 5940
 AGGTACATAA TGTTTGGGCC ACACATGCCT GTGTACCCAC AGACCCCAAC CCACAAGAAG
 5950 5960 5970 5980 5990 6000
 TAGTATTGGT AAATGTGACA GAAAATTTTA ACATGTGGAA AAATGACATG GTAGAACAGA
 6010 6020 6030 6040 6050 6060
 TGCATGAGGA TATAATCACT TTATGGGATC AAAGCCTAAA GCCATGTGTA AAATTAACCC
 6070 6080 6090 6100 6110 6120
 CACTCTGTGT TAGTTTAAAG TGCACATGTT TCCCGAATGC TACTAATACC AATAGTAGTA
 6130 6140 6150 6160 6170 6180

16 NOV 84- 29099

ATACCAATAG TAGTAGCGGG GAAATGATGA TGGAGAAAGG AGAGATAAAA AACTCTCTTT

D F A

20/26

6170 6200 6210 6220 6230 6240

TCAATATCAG CACAAGCATA AGAGGTAAGG TGCAGAAAGA ATATGCATTT TTTTATAAAC

6250 6260 6270 6280 6290 6300

TTGATATAAT ACCAATAGAT AATGATACTA CCAGCTATAC GTTGACAAGT TGTAACACCT

6310 6320 6330 6340 6350 6360

CAGTCATTAC ACAGGCCTGT CCAAAGGTAT CCTTTGAGCC AATTCCCATTA CATTATTCTG

6370 6380 6390 6400 6410 6420

CCCCGGCTGG TTTTGGGATT CTAAAATGTA ATAATAAGAC GTTCAATGGA ACAGGACCAT

6430 6440 6450 6460 6470 6480

GTACAAATGT CAGCACAGTA CAATGTACAC ATGGAATTAG GCCAGTAGTA TCAACTCAAC

6490 6500 6510 6520 6530 6540

TGCTGTTGAA TGGCAGTCTA GCAGAAGAAG AGGTAGTAAT TAGATCTGCC AATTTACACG

6550 6560 6570 6580 6590 6600

ACAATGCTAA AACCATAATA GTACAGCTGA ACCAATCTGT AGAAATTAAT TGTACAAGAC

6610 6620 6630 6640 6650 6660

CCAACAACAA TACAAGAAAA AGTATCCGTA TCCAGAGGGG ACCAGGGAGA GCATTTGTTA

6670 6680 6690 6700 6710 6720

CAATAGGAAA AATAGGAAAT ATGAGACAAG CACATTGTAA CATTAGTAGA GCAAAATGGA

6730 6740 6750 6760 6770 6780

ATGCCACTTT AAAACAGATA GCTAGCAAAT TAAGAGAACA ATTTGGAAAT AATAAAACAA

6790 6800 6810 6820 6830 6840

TAATCTTTAA GCAATCCTCA GGAGGGGACC CAGAAATTGT AACGCACAGT TTTAATTCTG

6850 6860 6870 6880 6890 6900

GAGGGGAATT TTTCTACTGT AATTCAACAC AACTGTTTAA TAGTACTTCC TTTAATAGTA

6910 6920 6930 6940 6950 6960

CTTGGAGTAC TGAAGGGTCA AATAACACTG AAGGAAGTGA CACAATCACA CTCCCATGCA

6970 6980 6990 7000 7010 7020

GAATAAAACA ATTTATAAAC ATGTGGCAGG AAGTAGGAAA AGCAATGTAT GCCCCTCCCA

7030 7040 7050 7060 7070 7080

TCAGCGGACA AATTAGATGT TCATCAAATA TTACAGGGCT CCTATTAACA AGAGATGCTG

7090 7100 7110 7120 7130 7140

GTAATAACAA CAATGGGTCC GAGATCTTCA GACCTGGAGG AGGAGATATC AGGGACAATT

7150 7160 7170 7180 7190 7200

GGAGAAGTGA ATTATATAAA TATAAAGTAG TAAAAATTGA ACCATTAGGA GTAGCACCCA

7210 7220 7230 7240 7250 7260

CCAAGGCAAA CAGAAGAGTG GTGCAGAGAG AAAAAAGAGC AGTGGGAATA GGAGCTTTGT

7270 7280 7290 7300 7310 7320

TCCTTGGGTT CTTGGGAGCA GCAGGAAGCA CTATGGGCGC ACGGTCAATG ACGCTGACCG

7330 7340 7350 7360 7370 7380

TACAGGCCAG ACAATTATTG TCTCGTATAG TGCAGCAGCA GAACAATTTG CTGAGGGCTA

7390 7400 7410 7420 7430 7440

10 NOV 84 - 28085

AGGGGCA ACAUCATCTG TTGCACTCA CAGTCTGGGG CATCAAGCAG CTCCAGGCAA

D F A

7450 7460 7470 7480 7490 7500
GAATCCTGGC TGTGGAAAGA TACCTAAAGG ATCAACAGCT CCTGGGGATT TGGGGTTGCT

7510 7520 7530 7540 7550 7560
CTGGAAAACCT CATTTCGACC ACTGCTGTGC CTTGGAATGC TAGTTGGAGT AATAAATCTC

7570 7580 7590 7600 7610 7620
TGGAACAGAT TTGGAATAAC ATGACCTGGA TGGAGTGGGA CAGAGAAATT AACAAATTACA

7630 7640 7650 7660 7670 7680
CAAGCTTAAT ACATTCTTA ATTGAAGAAT CGCAAAACCA GCAAGAAAAG AATGAACAAG

7690 7700 7710 7720 7730 7740
AATTATTGGA ATTAGATAAA TGGGCAAGTT TGTGGAATTG GTTTAACATA ACAAATTGGC

7750 7760 7770 7780 7790 7800
TGTGGTATAT AAAAATATTG ATAATGATAG TAGGAGGCTT GCTAGGTTTA AGAATAGTTT

7810 7820 7830 7840 7850 7860
TTGCTGTACT TTCTATAGTG AATAGAGTTA GGCAGGGATA TTCACCATT ACGTTTCAGA

7870 7880 7890 7900 7910 7920
CCCACCTCCC AACCCCGAGG GGACCCGACA GGCCCGAAGC AATAGAAGAA GAAGGTGGAG

7930 7940 7950 7960 7970 7980
AGAGAGACAG AGACAGATCC ATTCGATTAG TGAACGGATC CTTAGCACTT ATCTGGGACC

7990 8000 8010 8020 8030 8040
ATCTGGGGAG CTTTGTGCTT CTTGAGCTAC CACCGCTTGA GAGACTTACT CTTGATTGTA

8050 8060 8070 8080 8090 8100
ACGAGGATTG TGGAACTTCT GGGACGCAGG GGGTGGGAAG CCCTCAAATA TTGGTGGAAAT

8110 8120 8130 8140 8150 8160
CTCCTACAGT ATTGGAGTCA GGAAGTAAAG AATAGTGCTG TTAGCTTGCT CAATGCCACA

8170 8180 8190 8200 8210 8220
GCCATAGCAG TAGCTGAGGG GACAGATAGG GTTATAGAAG TAGTACAAGG AGCTTGTAGA

8230 8240 8250 8260 8270 8280
GCTATTGCGC ACATACCTAG AAGAATAAGA CAGCGCTTGG AAAGGATTTT GCTATAAGAT

8290 8300 8310 8320 8330 8340
GGGTGGCAAG TGGTCAAAAA GTAGTGTGCT TGGATGGCCT ACTGTAAGGG AAAGAATGAG

8350 8360 8370 8380 8390 8400
ACGAGCTGAG CCAGCAGCAG ATGGGGTGGG AGCAGCATCT CGAGACCTGG AAAACATGG

8410 8420 8430 8440 8450 8460
AGCAATCACA AGTAGCAATA CAGCAGCTAC CAATGCTGCT TGTGCCTGGC TAGAAGCACA

8470 8480 8490 8500 8510 8520
AGAGGAGGAG GAGGTGGGTT TTCCAGTCAC ACCTCAGGTA CCTTTAAGAC CAATGACTTA

8530 8540 8550 8560 8570 8580
CAAGGCAGCT GTAGATCTTA GCCACTTTTT AAAAGAAAAG GGGGGACTGG AAGGGCTAAT

8590 8600 8610 8620 8630 8640
TCACTCCCAA CGAAGACAAG ATATCCTTGA TCTGTGGATC TACCACACAC AAGGCTAGTT

8650 8660 8670 8680 8690 8700

1725

15 NOV 84- 230

6/26

D.F.A

CCCTGATTGG CAGAACTACA CACCAGGGCC AGGGGTGCGA TATCCACTGA CCTTTGGATC
8710 8720 8730 8740 8750 8760
GTGCTACAAG CTAGTACCAG TIGAGCCAGA TAAGGTAGAA GAGGCCAATA AAGGAGAGAA
8770 8780 8790 8800 8810 8820
CACCAGCTTG TTACACCCTG TGAGCCTGCA TGAATGGAT GACCCTGAGA GAGAAGTCTT
8830 8840 8850 8860 8870 8880
AGAGTGGAGG TTTGACAGCC GCCTAGCATT TCATCACGTG CCCCAGAGAGC TGCATCCGCA
8890 8900 8910 8920 8930 8940
GTACTTCAAG AACTGCTGAG ATCGAGCTTG CTACAAGGGA CTTTCCGCTG GGGACTTTCC
8950 8960 8970 8980 8990 9000
AGGGAGGCGT GGCCTGGCGG GAACTGGGGA GTGGCGAGCC CTCAGATGCT GCATATAACC
9010 9020 9030 9040 9050 9060
AGCTGCTTTT TGCCTGTACT GGGTCTCTCT GGTAGACCA GATTGAGCC TGGGAGCTCT
9070 9080 9090 9100 0 0
CTGGCTAACT AGGGAACCCA CTGCTTAAGC CTCAATAAAG CTT

Fig 2b